

Sınıflandırma Algoritmalarının Anestezi Türünün Tahminlemesi Üzerine Karşılaştırılması

Comparison of Classification Algorithms on The Estimation of Anesthesia Types

Gizem Dilan BOZTAŞ¹, Esra ÖZMEN², Ersin KARAMAN³

ÖZ

Veri madenciliği yöntemleri, reklam, ticaret, sağlık, eğlence, üretim, askeri gibi birçok alanda elde edilen verilerin sınıflandırılması, kümelenmesi ya da birbirleri ile ilişkilendirilmesi sonucunda gerektiğinde ileriye dönük tahminler oluşturmayı sağlamaktadır. Yapılacak tahminlemenin başarımlarına ise son derece önemlidir. Bundan dolayı veri madenciliği yöntemlerinde bulunan algoritmaların hangisinin daha iyi sonuç verdiği tahminleme açısından önem arz etmektedir. Bu çalışmada Erzurum Nenehatun Kadın Doğum Hastanesi'nde sezaryen yöntemiyle doğum yapan 300 gebeden elde edilen veri setine sınıflandırma algoritmalarından C4.5, regresyon ağacı, random forest, destek vektör makineleri ve k-en yakın komşu algoritmaları uygulanarak sezaryenle doğum yönteminde kullanılan anestezi türünün tahminleme başarımları karşılaştırılmıştır. Sınıflandırma algoritmalarının başarımlarını karşılaştırmaları yapılırken doğruluk, kesinlik, duyarlılık ve F-ölçütü göz önünde bulundurulmuştur. Çalışma sonucunda sezaryenle doğum yönteminde kullanılan anestezi türünün tahminleme başarımlarında doğruluk, duyarlılık ve F-ölçütüne göre en yüksek başarımları destek vektör makineleri sağlarken; kesinlik ölçütüne göre random forest algoritması en yüksek başarımlarını sağlamıştır.

Anahtar Kelimeler: Sınıflandırma algoritmaları, Tahminleme, Veri madenciliği

ABSTRACT

Data mining methods enable the creation of possible estimates as necessary as a result of the classification, clustering, or interrelation of data obtained in many areas such as advertising, trade, health, entertainment, production, military space. The performance rating of the forecast is extremely important. Therefore, it is important to predict which algorithms are better in data mining methods. In this study, data obtained from 300 pregnant women who delivered by cesarean section in Erzurum Nenehatun Gynecology Hospital were applied to C4.5, regression tree, random forest, support vector machines and k-nearest neighbor algorithms. Estimation success of the type of anesthesia used in cesarean delivery method was compared. Accuracy, precision, precision and F criteria were taken into account when performing performance comparisons of classification algorithms. At the end of the study, the accuracy, sensitivity and maximum performance support of the type of anesthesia used in the cesarean delivery method belongs to the support vector machines. According to the accuracy criterion, the highest performance ratio belongs to random forest algorithm.

Keywords: Classification algorithms, Estimation, Data mining

Bu çalışma 9-12 Ekim 2019 6. Uluslararası Yönetim Bilişim Sistemleri Konferansında (6th International Management Information Systems Conference) özet sözlü bildiri olarak sunulmuştur.

Çalışmadaki söz konusu veriler 532134 tez numarası ile ulusal tez merkezine kayıtlı olan tez çalışması kapsamında toplanmıştır.

¹ Öğr. Gör., Gizem Dilan BOZTAŞ, Yönetim Bilişim Sistemleri, Karadeniz Teknik Üniversitesi Uzaktan Eğitim Uygulama ve Araştırma Merkezi, gizem.boztas@hotmail.com, ORCID: 0000-0002-4593-032X

² Öğr. Gör. Dr., Esra ÖZMEN, Yönetim Bilişim Sistemleri, Ankara Hacı Bayram Veli Üniversitesi Uzaktan Eğitim Uygulama ve Araştırma Merkezi, esra.ozmen@hbv.edu.tr, ORCID: 0000-0001-6661-3473

³ Doç. Dr., Ersin KARAMAN, Yönetim Bilişim Sistemleri, Ankara Hacı Bayram Veli Üniversitesi İktisadi ve İdari Bilimler Fakültesi, ersin.karaman@hbv.edu.tr, ORCID: 0000-0002-6075-2779

İletişim / Corresponding Author: Esra ÖZMEN
e-posta/e-mail: esra.ozmen@hbv.edu.tr

Geliş Tarihi / Received: 28.09.2021

Kabul Tarihi/Accepted: 15.05.2021

GİRİŞ

Hayatımızın her anında ve yaşadığımız bütün alanlarda çeşitli veriler üretilmektedir. Fakat bu verilerin anlamlı hale gelmesi, doğru ve verimli hale dönüştürülmesi sonuç olarak doğru kararlar verebilmek için bilgi çıkarma işlemi veri madenciliği ile sağlanmaktadır.¹ Veri madenciliği yöntemleri, eğitimden sağlığa; ekonomiden eğlenceye uzanan çeşitli alanlarda kullanılmaktadır. Bu çalışma alanlarında kullanılan sınıflandırma algoritmaları için de alanyazında başarı oranlarının karşılaştırıldığı çeşitli çalışmalar yapılmaktadır. Örneğin Kale ve Yüksel (2020) e-posta önemliliğinin belirlenmesinde Decision Tree, Deep Learning, Gradient Boosted Tree, K-NN, Naive Bayes, Random Forest ve Logistic Regression sınıflandırma algoritmalarını ele almış ve bir başarı oranı karşılaştırması yapmıştır. Yapılan analizler sonucunda Gradient Boosted Tree algoritmasının en yüksek doğruluk değerine sahip olduğu görülmüştür.²

Demirci ve Karaatlı (2019) ise sınıflandırma algoritmalarını ülkelerin gelişmişlik düzeyinin tahmini için karşılaştırma yapmıştır. Değerlendirme sonucunda ele alınan yöntemlerin başarı oranı sırasıyla Attribute Selected Classifier algoritması, Bayesnet, JRIP ve J48 algoritmalarıdır.³

Veri madenciliği yöntemleri ile ilgili sağlık alanında da birçok çalışma yapılmıştır. Örneğin Daş ve Türkoğlu DNA dizilimlerinin sınıflandırmasında karar ağacı algoritmalarını karşılaştırdıkları çalışmada J48, LMT ve random forest algoritmalarını ele almışlardır. 4 bakteri türünün DNA dizilimleri incelenmiş ve tekrar eden nükleotit çiftleri frekansı bulunarak sınıflandırma gerçekleştirilmiştir. Yapılan çalışmada random forest algoritmasının, J48 ve LMT algoritmalarından sınıflandırma başarımının çok daha yüksek olduğu ve %100 doğruluk oranıyla sınıflandırma yaptığı sonucunda varılmıştır.⁴

Daş ve Türkoğlu'nun yaptıkları başka bir çalışmada ise DNA dizilimindeki nükleotit

çiftlerinin frekans değerlerine göre farklı sınıflandırma yöntemleri ile karşılaştırmışlardır. Bu çalışmada ise yapay sinir ağları (YSA), destek vektör makineleri ve k-en yakın komşu yöntemleri karşılaştırılmıştır. Sonuç olarak KNN yönteminin, YSA ve DVM yöntemlerinden başarımının daha yüksek olduğu görülmüştür.⁵

Veri madenciliği sınıflandırma algoritmalarının sınıflandırılması üzerine Coşkun ve Baykal'ın yaptığı bir çalışmada ise WEKA programı ile SEER veri kaynağı üzerinde J48, Naive Bayes, Lojistik Regresyon ve KStar algoritmaları kullanarak modeller oluşturulmuştur. Elde edilen modellerin tahminleme başarımları karşılaştırılarak kullanılan veri kaynağında hangi algoritmanın daha başarılı modeller oluşturduğu irdelenmiştir. Çalışma sonucunda doğruluk ve F-ölçütü açısından J48 algoritmasının, kesinlik açısından lojistik regresyon algoritmasının da yüksek değerlere sahip olduğu görülmüştür.⁶

Sağlık alanında yapılan başka bir çalışmada ise meme kanseri tespitinde sınıflandırma algoritmaları başarı oranları karşılaştırılmıştır. Çalışma sonucunda doğruluk oranı için Lojistik Regresyon (LR), En Yakın Komşu (KNN) Algoritması, Karar Ağacı (DecisionTree) yüksek başarımları gösterirken Destek Vektör Makineleri (SVM) diğer algoritmalar kadar yüksek bir doğruluk oranı göstermemiştir.⁷

Bu çalışmada ise günümüzde birçok sağlık uygulamasında başarılı sonuçlar üreten veri madenciliği yöntemlerinden sınıflandırma yöntemi ile sezaryen doğum yöntemiyle doğum yapmış 300 kadının ameliyatlarında tercih edilen anestezi yöntemlerini tahmin etme başarımları farklı ölçütlere göre karşılaştırılmıştır. Bu karşılaştırma bu alanda yapılacak yeni sınıflandırma çalışmalarına temel oluşturma açısından önem taşımaktadır.

Sezaryen yöntemi ile doğum, bazı durumlarda anne ve bebek açısından hayat kurtarıcı bir yöntem olabilmektedir. Sezaryen doğum sırasında kullanılan anestezi yönteminin hayati bir önem taşımasıyla

beraber doğum sonrası yaşam kalitesine de etkisinin doğrudanlığı düşünüldüğünde anestezi yönteminin doğru belirlenmesinin önemi çok daha iyi anlaşılmaktadır.

MATERYAL VE METOT

Veri madenciliği yöntemleri genel olarak öngörü ve tanımlayıcı yöntemler olmak üzere iki başlık altında toplanmaktadır. Öngörü yöntemleri ise sınıflandırma algoritmaları olarak incelenmektedir. Çalışma kapsamında sınıflandırma algoritmalarından; destek vektör makineleri, k-en yakın komşu, regresyon ağacı, C4.5 ve rastgele orman ağacı algoritmaları kullanılmıştır.

Destek vektör makineleri (DVM), öğrenme, sınıflandırma, tahminleme ve eldeki verilerden regresyon kuralları oluşturmak için kullanılan bir algoritmadır.⁸ DVM istatistiksel öğrenme teorisi üzerine inşa edilmiş ve 1960'lı yıllarda ilk olarak V. Vapnik tarafından ortaya çıkartılmış bir makine öğrenmesi yöntemi olarak bilinmektedir. DVM veri madenciliği alanında kullanılan değişkenler arasındaki bilinmeyen örüntülerin veri setlerindeki sınıflama problemlerinde faydalı olmaktadır. DVM en temel anlamda iki sınıfa ait verileri ayırabilecek en optimum düzlem ayırıcını belirlemektedir.⁹

K-en yakın komşu algoritması sınıflandırma problemlerine yönelik kullanılan denetimli öğrenme yöntemleri arasında yer almaktadır. Sınıflandırılacak verilerin öğrenme kümesindeki verilere benzerliklerinin hesaplanmasının ardından en yakın olduğu düşünülen k verinin ortalamasıyla belirlenmiş olan eşik değere göre sınıflandırma yapılır. Dikkat edilmesi gereken nokta ise her bir sınıf özelliğinin önceden net bir şekilde belirlenmiş olmasıdır.¹⁰ KNN algoritmasının oluşturulmasındaki etkili parametreler; k komşu sayısı, uzaklık ölçütü ve ağırlıklandırma parametreleridir.¹¹ Kısaca KNN var olan öğrenme verisine yeni bir veri geldiğinde mevcut veriler üzerinde sınıflandırma yapan eğitilmiş öğrenme algoritmasıdır.⁵

Kural tabanlı bir teknik olan sınıflandırma ve regresyon ağacı algoritmasında temel amaç bağımlı değişkene göre bağımsız değişkenler matrisini homojen olan alt sınıflara ayırmaktır. Sınıflar oluşturulurken veri dallanan bir ağaç yapısı ve hiyerarşik bir düzen alır. Bu algorithmada eğer bağımlı değişken kategorik ise yöntemin adı sınıflandırma ağacı, sürekli ise regresyon ağacı şeklinde isimlendirilir.¹² Regresyon ağaçlandırmada sınıflandırma ağacının tersine herhangi bir sınıf söz konusu değildir. Bu teknikte ayırımlar “artıkların karelerini azaltma algoritması” üzerinden gerçekleştirilir.¹³

C4.5 algoritması sayısal veriler ile karar ağaçlarının oluşturulmasına imkân vermektedir. Veriler içerisindeki çıkış değerini en fazla farklılaştıran öznitelik olarak seçilir ve bu özneliğin her değeri için farklı bir dal oluşturulmaktadır. C4.5 algoritması sayısal verileri kullandığı için bunları belirli bölümlere ayırmak zorluk oluşturabilir. Ancak bu bölme işi için en uygun t eşik değeri hesaplanarak belirli aralıklara bölme yapılabilir.¹⁴ Kısaca C4.5 algoritması sınıflandırmada en farklı özelliğe sahip değişkeni bulmaktadır. Söz konusu değişkeni bulmak için ise entropi kavramından faydalanmaktadır. Entropi, bir veri setindeki belirsizliği ve rastgeleliği ölçmek için kullanılmaktadır.¹⁵

Rastgele orman algoritması, ağaç tipinde olan sınıflandırma algoritmasıdır. Bu yöntemde istenilen ağaç sayısı kadar çalışma yapılabilmektedir.¹⁶ Bu algoritmalarda bir ağaç oluşturabilmek için kullanıcının 2 parametre tanımlaması gerekmektedir. Bu parametrelerden biri en iyi bölünmeyi belirlemek için değişkenlerin sayısı bir diğeri ise geliştirilecek ağaçların sayısıdır.¹⁷

Sınıflandırma model performans değerlendirmelerinde sıklıkla kullanılan

yöntemlerden biri olan karışıklık matrisi hata matrisi olarak da bilinir (Tablo 1). Modelin başarısı verilerin doğru veya yanlış sınıflandırılmasıyla ilgi olduğundan modelin performansını belirleyebilmek için hata matrisi üzerinden bazı performans ölçümleri yapılır.⁶ Birçok ölçüm çeşidi olmasına karşın doğruluk, hata oranı, kesinlik, duyarlılık, özgünlük ve F- ölçütü kullanılan temel ölçümlerdir.¹⁸

Tablo 1. Karışıklık Matrisi (Hata Matrisi)

		Gerçek Durum	
		Doğru (+)	Yanlış(-)
Tahmin Durumu	Doğru (+)	Gerçek Pozitif (a)	Yanlış Pozitif (b)
	Yanlış (-)	Yanlış Negatif (c)	Gerçek Negatif (d)

Temel ölçümlerin formülleri aşağıda belirtilmiştir.¹⁹

Doğruluk: En çok kullanılan ve en basit ölçüttür ($\text{Doğruluk} = a+d / a+b+c+d$).

Hata oranı: Doğruluğu 1'e tamamlayan değerdir. Yani yanlış sınıfa atanan verilerin toplamının toplam veri sayısına oranıdır ($\text{Hata oranı} = 1 - \text{doğruluk} \Rightarrow \text{hata oranı} = c+b / a+b+c+d$).

Kesinlik: Pozitif kestirim değeridir ($\text{Kesinlik} = a / a+b$).

Duyarlılık: Doğru pozitif oranıdır ($\text{Duyarlılık} = a/a+c$).

Kesinlik ve duyarlılık ölçümleri sadece olumlu gerçeklik ve öngörülere odaklandıklarından modelin olumsuz durumlarda işleyişi hakkında bilgi vermezler. Bu durum söz konusu iki değişkenin harmonik ortalaması olan F ölçümü için geçerlidir.²⁰

$$\text{Özgüllük} = d/d+b$$

$$F = 2 * \text{duyarlılık} * \text{kesinlik} / \text{duyarlılık} + \text{kesinlik}$$

Araştırmanın Örnekleme

Bu çalışmanın örneklemini Erzurum Nene Hatun Kadın Doğum Hastanesinde sezaryen

yöntemiyle doğum yapmış 300 kadın oluşturmaktadır.

Veri Toplama Araçları

Çalışmada kullanılan veri seti, uzman görüşü doğrultusunda hazırlanan 24 parametrelilik formun Erzurum Nenehatun kadın doğum hastanesinde sezaryen yöntemiyle doğum yapan 300 gebeye uygulanmasıyla oluşturulmuştur.

Veri Analizi

Çalışma için gerekli veri tabanı uzman görüşü ile oluşturulan hasta tanıma formu ile bilgiler toplanarak oluşturulmuş ve R yazılımı kullanılarak analiz edilmiştir.

Çalışma kapsamında toplanan veriler alan uzmanı görüşü doğrultusunda filtreleme yöntemlerinden olan Bilgi Kazanımı (Information Gain) ve Ki-Kare (Chi-Squared) metotları ile parametre seçimi yapılarak parametre sayısı 16'ya indirgenmiştir. Analiz 16 parametre üzerinden gerçekleştirilmiş olup söz konusu parametreleri 4 tane sayısal, 11 tane kategorik değişken oluşturmaktadır. Ayrıca analizlerde kullanılan veri setinde kayıp değer bulunmamakla birlikte 1 tane aykırı değer saptanmıştır. Aykırı değer yerine ilgili sütunun ortalaması dikkate alınmıştır. Betimsel istatistik için kullanılan yaş, boy ve kilo parametreleri normal dağılım gösterdiğinden standardize edilmemiştir. Kategorik veriler yerine sayısal değerler atanarak temsil edilmiştir. Kullanılan veri setinin bağımlı değişkenini %79,6 (239 kişi) spinal anestezi, %20,4 (61 kişi) genel anestezi oluşturmaktadır.

Oluşturulan 300 kişilik veri seti hold out yöntemi ile %80 eğitim verisi, %20 test verisi olarak ayrılmıştır. Uygulama için sınıflandırma algoritmalarından C4.5, regresyon ağacı, random forest, destek vektör makineleri ve k-en yakın komşu algoritmaları uygulanarak sezaryenle doğum yönteminde kullanılan anestezi türünün tahminleme başarıları doğruluk, kesinlik, duyarlılık ve F-ölçütüne göre karşılaştırılmıştır.

Araştırmanın Etik Yönü

Bu araştırmanın uygulanabilmesi için Atatürk Üniversitesi Tıp Fakültesi Klinik Araştırmalar Etik Kurulundan 15.02.2018 tarih ve 02/46 sayılı kararla onay alınmıştır. Ayrıca araştırmanın örneklem grubunda bulunan katılımcılardan gönüllülük beyanı hasta tanıma formu ile birlikte alınmıştır.

Araştırmanın Kısıtlılıkları

Çalışmada kullanılan verilerin sadece Erzurum ve çevresindeki hastalardan elde edilmesi yaş, boy, kilo, sezaryen sayısı, anestezi tercihleri gibi parametreler sonuçların genellenebilirliği açısından araştırmanın kısıdını oluşturmaktadır.

BULGULAR VE TARTIŞMA

Çalışma kapsamında ele alınan örneklem grubunun yaş, boy ve kilo parametrelerine ait betimsel istatistik verileri Tablo 2’de sunulmuştur.

Tablo 2. Örneklem Grubunun Yaş, Boy ve Kilo Parametrelerine Ait Betimsel İstatistik Verileri

Parametre	\bar{x}	ss (\pm)	Max.	Min	Medyan	N
Yaş	28,71	5,68	44	17	28,00	300
Boy	1,60	0,06	1,75	1,40	1,60	300
Kilo	78,07	12,09	113	54	77,00	300

Veri setindeki yaş, boy ve kilo parametrelerine ait betimsel istatistik değerleri Tablo 2’de gösterilmiştir. Tablo 2 incelendiğinde örneklem grubunu oluşturan 300 katılımcı için yaş niteliğinin ortalaması $28,71 \pm 5,68$ olarak, en büyük yaş değeri 44 olarak, en küçük yaş değeri 17 olarak tespit edilirken boy niteliğinin ortalaması $1,60 \pm 0,06$ cm olarak, en kısa boy oranı 1,40 cm olarak, en uzun boy oranı 1,75 cm olarak tespit edilmiştir. Kilo niteliğinin ise ortalaması $78,8 \pm 12,09$ kg olarak, en yüksek kilo değeri 113 olarak, en düşük kilo değeri 54 kg olarak tespit edilmiştir.

Kullanılan Algoritmaların Başarım Ölçütleri

C4.5 algoritması

Doğruluk	Kesinlik	Duyarlılık	F-Ölçütü
% 94,91	% 83,33	% 90,90	% 86,95

Regresyon

Doğruluk	Kesinlik	Duyarlılık	F-Ölçütü
% 89,83	% 83,33	% 71,42	% 76,92

Random forest algoritması

Doğruluk	Kesinlik	Duyarlılık	F-Ölçütü
% 96,61	% 100	% 83,33	% 90,90

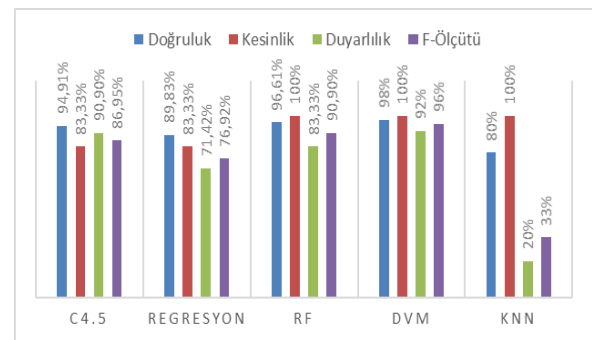
Destek vektör makinaları

Doğruluk	Kesinlik	Duyarlılık	F-Ölçütü
% 98	% 100	% 92	% 96

K- en yakın komşu algoritması

Doğruluk	Kesinlik	Duyarlılık	F-Ölçütü
% 80	% 100	% 20	% 33

K-en yakın komşu algoritması uygulanırken k değeri için birçok değer denenmiş fakat maksimum değerler $k=6$ olduğu durumda saptanmıştır. Kullanılan algoritmaların başarımlar ölçüt oranları Şekil 1’de sunulmuştur.



Şekil 1. Kullanılan Algoritmaların Başarım Ölçütü Oranları

Şekil 1’den anlaşılacağı üzere doğruluk ölçütü bağlamında Destek Vektör Makinaları en yüksek başarımları göstermiştir. Bu sonuçtan farklı olarak Coşkun ve Baykal (2011) kendi veri setlerinde J48 algoritmasının doğruluk ölçütünü daha başarılı bulurken Öztad (2020) ise Lojistik Regresyon doğruluk başarımlarını daha yüksek olduğunu ifade etmiştir.^{6,7}

Kesinlik ölçütüne bakıldığında en yüksek başarıyı random forest algoritması, destek vektör makinaları ve kNN algoritması sağlamaktadır. Fakat bu sonuç Coşkun ve Baykal (2011)'in Lojistik Regresyon algoritmasının kesinlik ölçütüne göre daha fazla başarı sağladığı sonucuyla farklılık göstermektedir.⁶ Bu farklılığında veri setlerinden kaynaklandığı söylenebilir.

F-ölçütü, doğruluk, kesinlik ve duyarlılık başarımları ölçütleri ele alındığında hepsi için en yüksek başarımları gösteren algoritmanın Destek Vektör Makinaları olduğu görülmektedir. Bu sonuçtan farklı olarak

Kale ve Yüksel (2020) Gradient Boosted Tree algoritmasının, Demirci ve Karaatlı (2019) Attribute Selected Classifier algoritmasını, Daş ve Türkoğlu (2014) En Yakın Komşu algoritmasının başarımlarını daha yüksek bulmuştur.²⁻⁴ Sonuçlardaki bu farklılığın çalışma alanı, veri türleri, başarımları ölçütleri ve kullanılan algoritma farklılıklarıyla ilgili olduğu düşünülmektedir.

SONUÇ VE ÖNERİLER

Çalışma sonucunda doğruluk ölçütüne göre, destek vektör makinalarının %98 doğruluk oranıyla en yüksek performansı gösterdiği görülmektedir. Bu bağlamda destek vektör makinalarını sırasıyla random forest algoritması, C 4.5 algoritması, regresyon ve kNN algoritması takip etmektedir.

Kesinlik ölçütüne göre random forest algoritması, destek vektör makinaları ve kNN algoritması aynı performansı gösterirken bunları C 4.5 algoritması ve regresyon yine eşit performans göstererek takip etmektedir. Duyarlılık ölçütüne göre en yüksek performansı gösterme sıralaması ise, destek vektör makinaları, C4.5 algoritması, random forest algoritması, regresyon ve kNN algoritması şeklindedir. Kesinlik ve duyarlılık ölçütlerinin harmonik ortalaması olan F-ölçütüne bakıldığında sıralamanın destek vektör makinaları, random forest

algoritması, C4.5 algoritması, regresyon ve kNN şeklinde olduğu belirlenmiştir.

Bu çalışmada söz konusu algoritmaların performans değerlendirmeleri çeşitli ölçütlere göre karşılaştırılmıştır. Literatürde sağlık bilimleri alanında yapılmış benzer çalışmalar bulunsa da kullanılan veri setinin hem numerik hem de kategorik verilerden oluşmasının çalışmayı benzerlerinden ayırdığı düşünülmektedir. Sağlık alanında veri bilimi çalışmalarının ilerleyen zamanlarda artacağı öngörüsünden hareketle, bu çalışmaların farklı veri setleri ve yöntemlerle denenerek alan yazına katkıda bulunulabileceği düşünülmektedir. Ayrıca hem yapay zeka teknikleri hem de veri madenciliği kapsamında kümeleme ve birliktelik kurallarına yönelik algoritmalarının sağlık alanına has veri karakteristikleri ile değerlendirilmesi gelecek çalışmalara temel oluşturacaktır.

KAYNAKLAR

1. Argüden, Y. ve Erşahin, B. (2008). "Veri Madenciliği: Veriden Bilgiye" Masrafdan Değere". İstanbul: Alkim Kağıt Sanayi ve Ticaret A.Ş.
2. Kale, B. ve Yüksel G. (2020). "Veri Madenciliği Sınıflandırma Algoritmaları İle E-posta Önemliliğinin Belirlenmesi". Ç.Ü Fen ve Mühendislik Bilimleri Dergisi, 39 (9), 109-116.
3. Demirci, E. ve Karaatlı, M. (2019). "Ülkelerin Gelişmişlik Seviyelerinin Tahmininde Kullanılan Sınıflandırma Algoritmalarının Karşılaştırılması". Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 24 (3), 703-714.
4. Daş, B. ve Türkoğlu, İ. (2014). "DNA Dizilimlerinin Sınıflandırılmasında Karar Ağacı Algoritmalarının Karşılaştırılması". Elektrik - Elektronik - Bilgisayar ve Biyomedikal Mühendisliği Sempozyumu (ELECO 2014), (s.381-383). Bursa.
5. Daş, B. ve Türkoğlu, İ. (2014). "DNA Dizilimlerindeki Nükleotit Çiftlerinin Frekans Değerlerine Göre Farklı Sınıflandırma Yöntemleri ile Karşılaştırılması". Tıp Teknolojileri Ulusal Kongresi, (s.191-194). Kapadokya.
6. Coşkun, C. ve Baykal, A. (2011). "Veri Madenciliğinde Sınıflandırma Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması". Akademik Bilişim'11 - XIII. Akademik Bilişim Konferansı Bildirileri, (s.51-58), Malatya.
7. Öztad, E. (2020). "Meme Kanseri Tespitinde Sınıflandırma ve Sinir Ağları Yöntemlerinin Karşılaştırılması". İleri Mühendislik Çalışmaları ve Teknolojileri Dergisi, 1 (1), 49-54.
8. Çomak, E. (2004). "Destek Vektör Makineleri Çoklu Sınıf Problemleri İçin Çözüm Önerileri". Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
9. Yılmaz, Ö, Ankaralı, H, Aydın, D. ve Saraçlı, Ö. (2013). "Tıbbi Tahminde Alternatif Bir Yaklaşım: Destek Vektör Makineleri". Türkiye Klinikleri Biyoistatistik, 5(1), 19.
10. Çalışkan, S. ve Soğukpınar, İ. (2008). "KNN: K-Means ve K-En Yakın Komşu Yöntemleri İle Ağlarda Nüfuz Tespiti". 2. Ağ ve Bilgi Güvenliği Sempozyumu, (120-124).
11. Taşçı, E. ve Onan, A. (2016). "K-En Yakın Komşu Algoritması Parametrelerinin Sınıflandırma Performansı Üzerine Etkisinin İncelenmesi". Akademik Bilişim.
12. Özkan, K. (2012). "Sınıflandırma ve Regresyon Ağacı Tekniği (SRAT) ile Ekolojik Verinin Modellenmesi". Turkish Journal of Forestry, 13(1), 1-4.
13. Breiman, L, Friedman, J.H, Olshen, R.A. and Stone, C.J. (1984). "Classification and Regression Trees. Monterey", CA: Wadsworth & Brooks/Cole Advanced Books & Software.
14. Kökver, Y, Barışçı, N, Çiftçi, A. ve Ekmekçi, Y. (2014). "Hipertansiyona Etki Eden Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi". E-Journal of New World Sciences Academy, 15-25.
15. Danacı, M, Çelik, M. ve Akkaya, A.E. (2010). "Veri Madenciliği Yöntemleri Kullanılarak Meme Kanseri Hücrelerinin Tahmin ve Teşhisi". Akıllı sistemlerde Yenilikler ve Uygulamaları Sempozyumu ASYU'2010 (s.21-24).
16. Ok, A, Akar, Ö. ve Güngör, O. (2019). "Rastgele Orman Sınıflandırma Yöntemi Yardımıyla Tarım Alanlarındaki Ürün Çeşitliliğinin Sınıflandırılması". Erişim adresi: <https://www.researchgate.net> (Erişim tarihi: 14 Ocak 2019).
17. Pal, M. (2005). "Random Forest Classifier for Remote Sensing Classification". International Journal of Remote Sensing, 26 (1), 217-222.
18. Selçukcan-Erol, Ç. (2016). "Sağlık Bilimlerinde R ile Veri Madenciliği". In: M.E. Balaban, E. Kartal (Ed.). R ile Veri Madenciliği Uygulamaları (s.26). İstanbul: Çağlayan Kitabevi.
19. Altunkaynak, B. (2017). "Veri Madenciliği Yöntemleri ve R Uygulamaları". Ankara: Seçkin Yayıncılık.
20. Powers, M.W. (2011). "Evaluation: From "Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation". Journal of Machine Learning Technologies, 2 (1), 37-63.